

# What is needed to exploit knowledge of primary transmissions?

Pulkit Grover

Wireless Foundations, EECS Department  
University of California at Berkeley  
Email: pulkit @ eecs.berkeley.edu

Anant Sahai

Wireless Foundations, EECS Department  
University of California at Berkeley  
Email: sahai @ eecs.berkeley.edu

**Abstract**—Recently, Tarokh and others have raised the possibility that a cognitive radio might know the interference signal being transmitted by a strong primary user in a non-causal way, and use this knowledge to increase its data rates. However, there is a subtle difference between knowing the signal transmitted by the primary and the actual interference at our receiver since there is a wireless channel between these two points. We show that even an unknown phase results in a substantial decrease in the data rates that can be achieved, and thus there is a need to feedback interference channel estimates to the cognitive transmitter. We then consider the case of fading channels. We derive an upper bound on the rate for given outage error probability for faded dirt. We give a scheme that uses appropriate “training” to obtain such estimates and quantify this scheme’s required overhead as a function of the relevant coherence time and interference power.

## I. INTRODUCTION

Knowledge of different aspects of a wireless channel can be exploited in various ways to increase the achievable data rates. For example, consider an OFDM transmission when the channel has frequency selective fading. The receiver needs to know the channel in order to do equalization and thereby interpret the information on the various subcarriers. To support this equalization, the transmitter helps by dedicating some of its energy to transmitting known pilot tones [1]. As the bandwidth gets large and the number of fading parameters increases, the overhead required to learn these parameters at the receiver increases and it has been shown [2], [3] that the optimal signaling in wideband communications is ‘peaky’ — concentrating most of its energy in a few time/frequency slots.

The receiver clearly needs to implicitly or explicitly learn the wireless channel, but in many cases, the system can benefit from having the transmitter exploit this knowledge as well. If the channel is frequency selective, then the transmitter can use the simple water-pouring

scheme [1] to allocate its power across subchannels to achieve optimal rates. [4] extends this insight to practical settings where the signal constellations are constrained. If the fading is not frequency selective, then for a single-input single-output channel, transmitter knowledge does not really impact waveform design.

In practice, channel knowledge is never perfect. Channel uncertainty at the receiver comes from the finite underlying coherence time of the time-varying wireless channel. In wireless systems that time-share a single frequency between forward and backward links, wireless reciprocity induces a similar channel uncertainty at the transmitter. More typically, the limitation in channel knowledge at the transmitter comes from the quality of the feedback link over which the channel information is sent back. Such uncertainty in knowledge of the channel is known to result in appreciable degradation in performance[5], [6], [7]. Even for flat fading, [5] shows that for the MIMO broadcast channel, fading uncertainty reduces the throughput substantially at high SNR.

This paper explores the impact of channel uncertainty in the cognitive radio context. Generally, a cognitive radio senses its environment, learns from it, and opportunistically uses the channel resources. While in many cases, the focus is on finding empty bands in which to transmit, [8], [9], [10] suggest a radically different perspective. They suggest that very strong interference from a primary user can be exploited by a communication system to increase its achievable rate beyond what is possible by simply treating the interference as noise.

If the receiver perfectly knows the interfering signal, it can simply subtract it off from the received signal and thereby pretend that there is no interference. It makes no difference if the knowledge is delayed, instantaneous, or non-causal. The idea that noncausal knowledge could be useful at the *transmitter* is attributed to Gelfand and Pinsker [11]. In a surprising result, Costa [12] proved

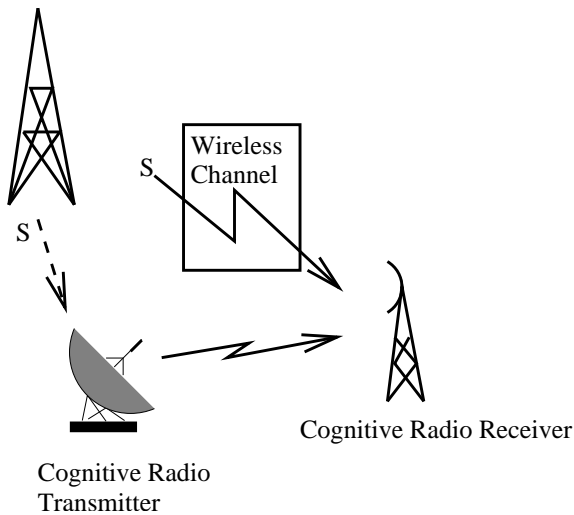


Fig. 1. A cognitive radio cognizant of signal transmitter by the primary. For simplicity, we assume that the interferer's channel results in a phase shift,  $e^{j\theta}$ .

that the capacity of a channel with additive Gaussian interference (known non-causally at the transmitter) is same as that with no interference at all. This is called “dirty paper coding” and it forms the basis for the schemes in [8], [9], [10] where it is assumed that the transmitter is able to decode what the primary is sending before it begins its own transmissions.

Fig. 1 illustrates the setup. Once again, it is natural to consider uncertainty in the knowledge of fading. After all, there is a difference between knowing the transmitted primary signal and knowing the interference caused by this signal at our own receiver. We consider the case of flat fading and concentrate on the unknown phase shift for simplicity. We pose the following questions:

- How important is it for the secondary *transmitter* to have precise interference phase knowledge?
- How does the fade amplitude affect the achievable rates?
- Does feedback help and if so, how should we use it?

The problem of uncertainty in knowledge of the interference at the transmitter has been addressed in [13][14]. [13] models interference uncertainty by allowing the channel interference vector to take one of two possible values, both known at the transmitter only. The two possible interference vectors are drawn independently with Gaussian distributions. The transmitter does not know which particular value the interference assumes. [13] shows that situation is equivalent to a broadcast channel model with a common message, where the

transmitter knows the interference vectors to the two receivers. They are able to show an explicit rate penalty for not knowing the interference exactly. For the scenario in Fig. 1, however, the model in [13] is not realistic: the possible realizations of the interference vectors are far from independent. It is unclear from [13] whether phase-uncertainty alone would impose a significant rate penalty.

[14] considers the general problem of uncertainty in knowledge of interference, modeling it as a compound Gelfand-Pinsker problem. This model does capture the scenario in Fig. 1 and gives the following lower bound  $C_l$  and upper bound  $C_u$  on the capacity  $C$ .

$$C_l = \sup_{P_{U|X,S,W}, P_{X|S,W}, P_W} \inf_{\beta \in \mathcal{C}} [I^\beta(U; Y|W) - I(U; S|W)] \quad (1)$$

$$C_u = \sup_{P_{X|S,W}, P_W} \inf_{\beta \in \mathcal{C}} \left[ \sup_{P_{U|X,S,W}} I^\beta(U; Y|W) - I(U; S|W) \right] \quad (2)$$

where  $\beta \in \mathcal{C}$  is a parameter characterizing the uncertainty in the knowledge of state  $S$ . Notice that both bounds are expressed as supremums over the distribution of an auxiliary random variable  $U$ . Evaluated for any particular  $U$ , (1) gives a lower bound to the capacity. However, any particular  $U$  *does not* give a valid upper bound in (2). That requires taking a supremum over all possible  $U$ 's, and it is not clear from [14] what the appropriate choice of  $U$  is. Without such guidance, the upper bound (2) is not computable for even a simple wireless channel model.

In addition to the bounds, a scheme is suggested in [14] for flat fading channels. The scheme uses Costa's dirty-paper strategy, where the auxiliary random variable  $U$  is defined to be equal to  $X + \alpha S$ . A seemingly reasonable value of  $\alpha$  is chosen, according to the distributions of the fade parameters. Unfortunately, the constant  $\alpha$  becomes zero if the fading coefficient has zero mean. Since Rician fading has a non-zero mean, [14] concentrates on that case. For simple phase uncertainty, the strategy is not optimal since  $\alpha = 0$  implies ignoring the interference knowledge at the transmitter. In Section V we show a scheme (inspired by [13]) that performs better at low SINR.

Sacrificing generality, we obtain stronger bounds than [14] for particular case of phase uncertainty. Section IV shows that the lack of phase knowledge can substantially reduce the attainable rates. This is done by modifying the bounding strategy in [13]. Section VI considers the case when low-rate feedback is available from the receiver. We show that if the transmitter encodes certain “training

data”, an asymptotically vanishing rate-loss is incurred by having to learn the phase to some fidelity and and communicate it back to the transmitter. For the residual uncertainty in phase, we give a modified dirty-paper coding scheme that achieves the capacity for perfect phase knowledge in the limit of uncertainty going to zero.

We then use the de consider the problem of fading uncertainty regarding the primary’s transmission, where the magnitude of the fade is also uncertain. In the classical point-to-point communication problem, a deep channel fade necessarily introduces errors, even with complete knowledge of the fading coefficient. Therefore, the problem of interest there is finding the achievable rates for given outage probability. However, here we consider fading only for the primary’s signal, which is really the interference to the secondary. With complete knowledge of the fading coefficient, the problem reduces to Costa’s dirty-paper coding problem [12] and there is no channel outage due to known interference fading.

Consider Rayleigh fading of the primary’s signal. Suppose our scheme ignores the knowledge of the primary’s transmission. The magnitude of the fading coefficient of the primary’s transmission can be arbitrarily large. Hence, the interference power can be arbitrarily large too. The SINR, and hence rate, achieved by this scheme would therefore be zero! This suggests considering achievable rates for a given outage probability. On the other hand, the trivial upper bound, corresponding to complete knowledge of the fading coefficient, suggests that outage probability does not change the rate significantly. Therefore, it is not clear if allowing for some outage probability would make a significant difference to the achievable rate.

From a system design perspective, we conclude that in designing systems that exploit non-causal knowledge of the interference, feedback has an important role to play. Furthermore, the natural “broadcast nature” of wireless transmissions will be destroyed since the transmitted data is targeted at a particular phase of the interference. This means that schemes like [15] will not function in this environment.

## II. PROBLEM STATEMENT

Consider the following complex channel model.

$$Y_i = X_i + h_i S_i + Z_i \quad (3)$$

where  $Y_i$  denotes the received signal at the secondary receiver at the time instant  $i$ ,  $X_i$  denotes the signal transmitted by the secondary transmitter,  $S_i$  denotes the

signal of the primary,  $h_i$  is the channel fade coefficient, and  $Z_i$  is AWGN. We are interested in maximizing the rate for the secondary.

For simplicity, we first assume that the fade amplitude,  $|h_i| = 1$ , and the phase of  $h_i$  is constant  $\theta$ , independent of  $i$ . Also, we assume block encoding and decoding, with block-length  $n$ . Therefore, the channel model is

$$\mathbf{Y}^n = \mathbf{X}^n + \mathbf{S}^n e^{j\theta} + \mathbf{Z}^n \quad (4)$$

where the superscript  $n$  denotes an  $n$ -length vector. The interference is modeled as (complex) Gaussian i.i.d at each instant, with variance  $Q$  and mean zero. The signal  $X$  is power constrained, in that the average power should not exceed  $P$ . The noise is AWGN, and without loss of generality, its variance is assumed to be 1. Under this model, we seek to find the maximum achievable rate under no knowledge of the phase  $\theta$  at the secondary transmitter.

We then consider the case when the fade amplitude  $|h_i| = \gamma$  and the phase is again unknown. The channel model now is

$$\mathbf{Y}^n = \mathbf{X}^n + \mathbf{S}^n e^{j\theta} + \mathbf{Z}^n \quad (5)$$

We want to find the achievable rate region for unknown  $\gamma$  and  $\theta$ .

In what follows,  $W$  denotes the message to be communicated by the cognitive radio.  $W$  is assumed to be chosen uniformly from the set of  $2^{nR}$  messages.

## III. UPPER BOUND ON THE COMMUNICATION RATE UNDER PHASE UNCERTAINTY OF THE INTERFERENCE

In this section, we give an upper bound on the communication rate assuming the phase of the interference is unknown at the transmitter.

In [13], the authors found bounds on the rate assuming that interference vector lies in a two-point set known to the transmitter. They observed that the problem of unknown interference is the same as that for communicating the same message simultaneously to two receivers, where the receivers face different interference vectors. We build on their idea.

Consider the case when the transmitter is unsure whether the phase at the receiver is 0 or  $\phi$ . This uncertainty is certainly no worse than no knowledge of the phase. Hence, any upper bound for this two-point uncertainty set is an upper bound for our case of no phase knowledge.

Suppose the uncertainty is between  $\mathbf{S}^n$  and  $\mathbf{S}^n e^{j\phi}$ , for some fixed  $\phi \leq \pi$ . Denote by  $\mathbf{Y}_1^n$  the output

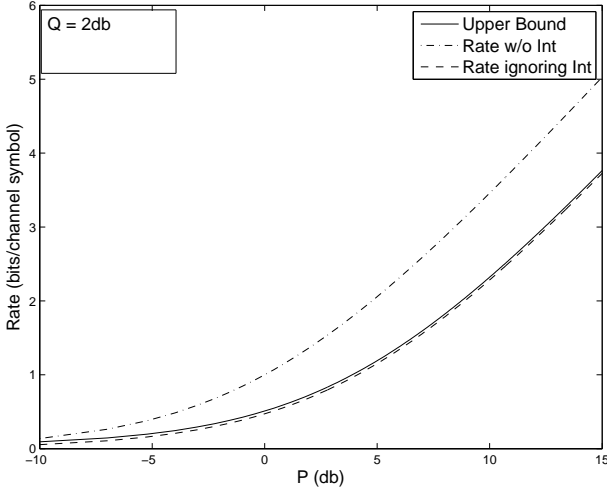


Fig. 2. Bounds on the rate for interference power  $Q = 2\text{db}$ . The bound shows that at moderate SIR, ignoring interference and treating it as noise is close to optimal if we do not know the phase of the interference at the receiver.

corresponding to the first user and  $\mathbf{Y}_2^n$  as the output corresponding to the second user. Then,

$$\begin{aligned} \mathbf{Y}_1^n &= \mathbf{X}^n + \mathbf{S}^n + \mathbf{Z}_1^n \\ \mathbf{Y}_2^n &= \mathbf{X}^n + \mathbf{S}^n e^{j\phi} + \mathbf{Z}_2^n \end{aligned} \quad (6)$$

For this model, the following upper bound holds:

*Theorem 1:* The rate for reliable communication for the channel model described in Section II is bounded by

$$R \leq \frac{1}{2} \log \left[ \frac{(P + Q + 1)^2}{4Q} \right] \quad (7)$$

*Proof:* See Appendix I.

As an aside, we note that the bounding technique here can also be used to find bounds on the rate for uncorrelated interference vectors of [13]. The bounds obtained are tighter than the bounds in [13] at high SIR.

The bound for rate is plotted as a function of transmit power  $P$  for interference power  $Q = 2\text{db}$  in Fig. 2, for  $Q = 5\text{db}$  in Fig. 3 and for  $Q = 15\text{db}$  in Fig. 4. At low  $Q \approx 2\text{db}$ , the bound is close to the rate achieved by ignoring the interference knowledge completely and treating it as noise. This shows that the bound is quite tight at low  $Q$  and that missing a single bit of phase-knowledge, makes the rest of the transmitter's knowledge of the interference almost useless! For increasing  $Q$ , the bound is seen to get looser. In fact, it is evident from Fig. 4 that the bound is extremely loose at low  $P$  and large  $Q$  since it goes above the 'no-interference' bound.

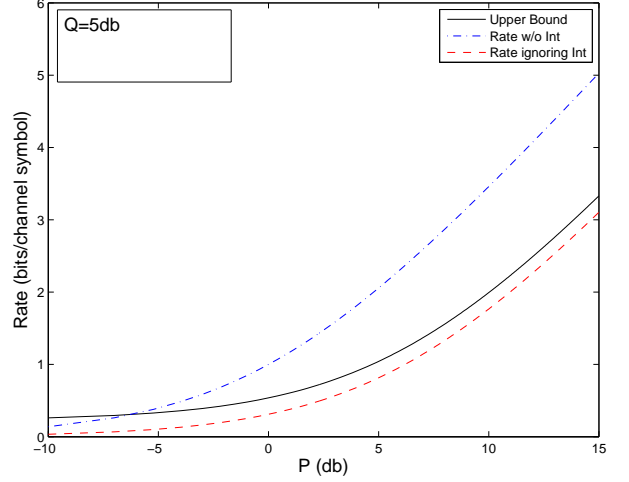


Fig. 3. Bounds on the rate for  $Q=5\text{db}$ . At low SIR, our new bound becomes very loose — going above the rate with zero interference.

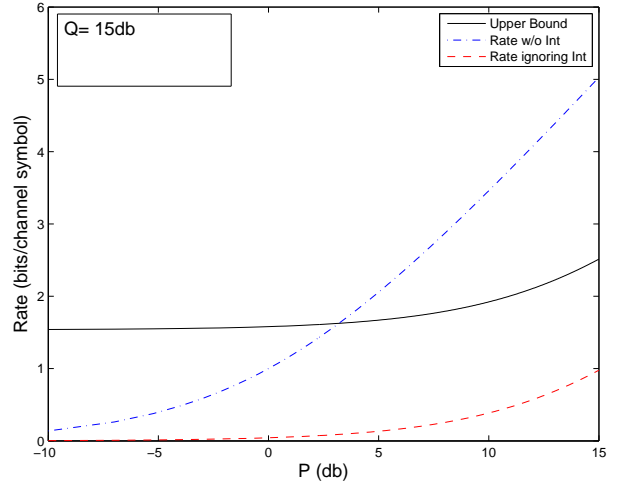


Fig. 4. The plot shows the bounds on the rate for  $Q = 15\text{db}$ . Our upper bound is quite loose at low signal powers. Even at high signal powers, it is unclear if the bound is tight because we have no achievable schemes that approach this performance.

#### IV. UPPER BOUND ON THE OUTAGE CAPACITY UNDER FADE UNCERTAINTY OF THE INTERFERENCE

In this section, we give an upper bound on the communication rate assuming the fading coefficient of the interference is unknown at the transmitter.

Assume first that the fade amplitude  $\gamma$  is fixed. We obtain computable upper bounds on the rate for given  $\gamma$ , using Theorem 1. Suppose the transmitter commits to a transmission rate  $R$ . We find the probability that the upper bound corresponding to the fade amplitude is

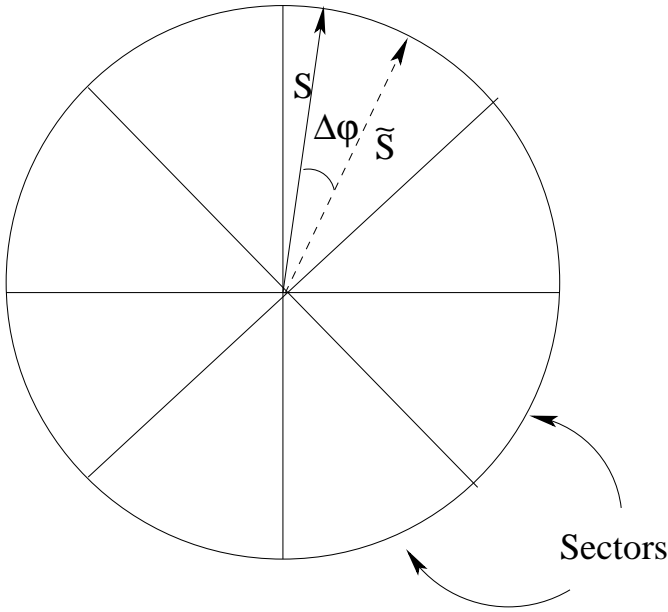


Fig. 5. The plot shows the idea behind the sectoring strategy. The phase-space is divided into  $k = 8$  sectors, and the transmitter dirty-paper codes according to the central vector in each sector,  $\tilde{S}$ . The actual interference is  $S$ , and the phase difference is  $\Delta\phi$

below this rate. This gives a lower bound on the outage probability.

We denote the capacity of the channel for given fixed  $\gamma$  by  $C(\gamma)$ .

For this model, the following upper bound holds:

*Corollary 1:* The rate for reliable communication for the channel model described in Section II with fixed fading coefficient  $\gamma$  is bounded by<sup>1</sup>

$$R \leq \frac{1}{2} \log \left[ \frac{(P + \gamma^2 Q + 1)^2}{4\gamma^2 Q} \right] \quad (8)$$

The bound is denoted by  $C_u(\gamma)$ .

*Proof:* Follows from Theorem 1.  $\square$

*Corollary 2:* Assuming that the transmitter now commits to a rate  $R$ , the outage probability is bounded by

$$p_{\text{out}} \geq \Pr(C_u(\gamma) < R) \quad (9)$$

*Proof:* The bound follows immediately from the fact that  $p_{\text{out}} \geq \Pr(C(\gamma) < R)$  and  $C(\gamma) \leq C_u(\gamma)$ .  $\square$

The bound on outage probability depends on the fade distribution through the random variable  $C_u(\gamma)$ .

The bound on the rate is plotted as a function of the

<sup>1</sup>As an aside, we note that the bounding technique here can also be used to find bounds on the rate for uncorrelated interference vectors of [13]. The bounds obtained are tighter than the bounds in [13] at high SIR.

fade amplitude  $\gamma$  for  $P = Q = 10$  db in Fig. 6. The lower figure in Fig. 6 is the pdf of Rayleigh distribution for the parameter  $\sigma^2 = 1$ . For fixed rate  $R = 2$  bits/symbol, the probability  $\Pr(R > C_u(\gamma))$  is the area of the shaded region in Fig. 6. This is a lower bound on the outage probability. Alternatively, for given outage probability, the curve gives an upper bound on the achievable rate.

The bound on the rate in Fig. 6 has a surprising feature: for large  $\gamma$ , it increases on increasing  $\gamma$ . We think this is an artifact of the bounding technique used, and is not fundamental to the problem at hand.

The bound thus obtained is plotted as a function of outage probability for various Rayleigh parameters in Fig. 7, and as a function of Rayleigh parameter for outage probability of 0.1 in Fig. 8. Since the bound in Fig. 6 is loose for large  $\gamma$ , the same is expected for the bound in Fig. 8 for large  $\sigma^2$ .

## V. A SCHEME FOR ACHIEVING IMPROVED RATES WITHOUT PHASE KNOWLEDGE

The bound above raises a natural question. Is it at all useful to have the interference knowledge, without the phase? The curves in Fig. 3 suggest that our bound is loose at low SIR, therefore, it is likely that the greatest advantage lies there.

At low SIR, a natural strategy is to break the interference uncertainty ‘circle’ into a few sectors. Following [13], the transmitter time shares between coding strategies that dirty paper code according the ‘central’ interference vector in each sector. We call this strategy ‘sectoring’. The strategy is illustrated in 5. To analyze such a strategy, however, requires us to find the impact of small phase uncertainty on the achievable rates. The problem is non-trivial because the residual interference is not independent of the interference vector, and hence not independent of the codeword to be detected.

### A. Achievable rates with small uncertainty in the phase

Assuming that the uncertainty in phase is a small value  $\Delta\phi$ , we investigate achievable rates for this reduced uncertainty. We wish to dirty paper code with respect to the ‘central’ interference vector in the uncertainty set, and treat the ‘residual’ uncertainty as noise.

Assume that the the phase  $\phi \in [-\Delta\phi, \Delta\phi]$  for some  $0 < \Delta\phi < \frac{\pi}{2}$ . The transmitter chooses the codeword  $\mathbf{U}^n$  according to the usual rule  $\mathbf{U}^n = \mathbf{X}^n + \alpha \mathbf{S}^n$ , where  $\mathbf{X}^n$  is chosen independent of  $\mathbf{S}^n$  as in [12]. The value of  $\alpha$  depends on  $\Delta\phi$ , as well as the SNR. The output  $\mathbf{Y}^n$  is

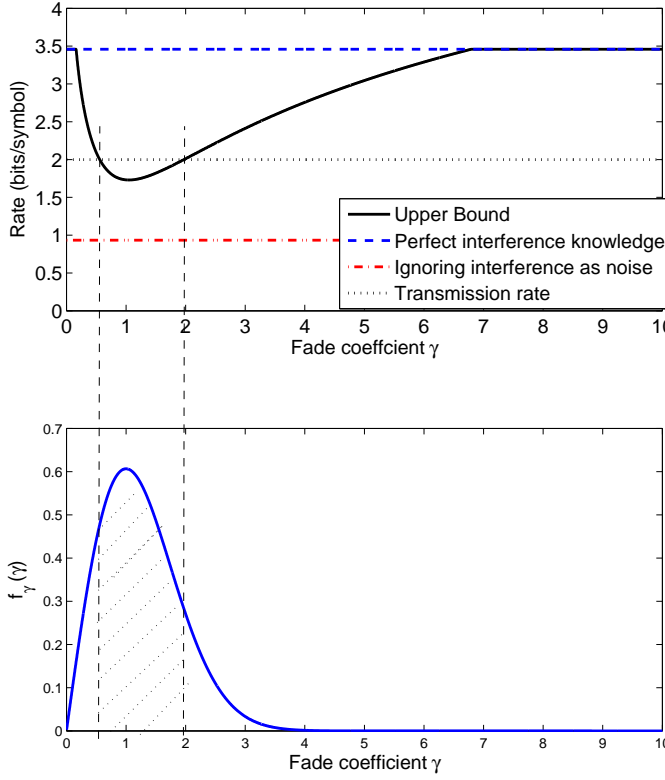


Fig. 6. The upper figure gives the upper bound on the rate for  $P = Q = 10\text{db}$  and varying fade coefficient  $\gamma$ . Also plotted are the rates with perfect knowledge of the interference (the classical dirty-paper coding) and the rate ignoring the knowledge of the interference. The lower figure is the pdf of Rayleigh distribution for  $\sigma^2 = 1$ . Assuming the transmission rate to be  $R = 2$  bits/symbol, the calculation of lower bound on outage probability is shown. The lower bound is the area of the shaded region in the lower figure. The slowest increase in the rate with increase in outage probability would be when the peak of Rayleigh distribution lies near the valley of the bound. A surprising aspect of the upper bound is that for large  $\gamma$ , it is an increasing function of  $\gamma$ . We think this is an artifact of our bounding technique, and is not the actual behavior of the capacity.

given by

$$\mathbf{Y}^n = \mathbf{X}^n + \mathbf{S}^n e^{j\phi} + \mathbf{Z}^n \quad (10)$$

$$= \mathbf{U}^n - \alpha \mathbf{S}^n + \mathbf{S}^n e^{j\phi} + \mathbf{Z}^n \quad (11)$$

Since  $\mathbf{X}^n$  and  $\mathbf{S}^n$  are independent,  $\mathbf{U}^n$  is jointly Gaussian with  $\mathbf{S}^n$ . Therefore,  $\mathbf{U}^n$  and  $\mathbf{Y}^n$  are also jointly Gaussian<sup>2</sup>.

We prove the following:

*Theorem 2:* For a phase uncertainty of  $\Delta\phi$ , an achiev-

<sup>2</sup>Any linear combination of these two random variables is jointly Gaussian

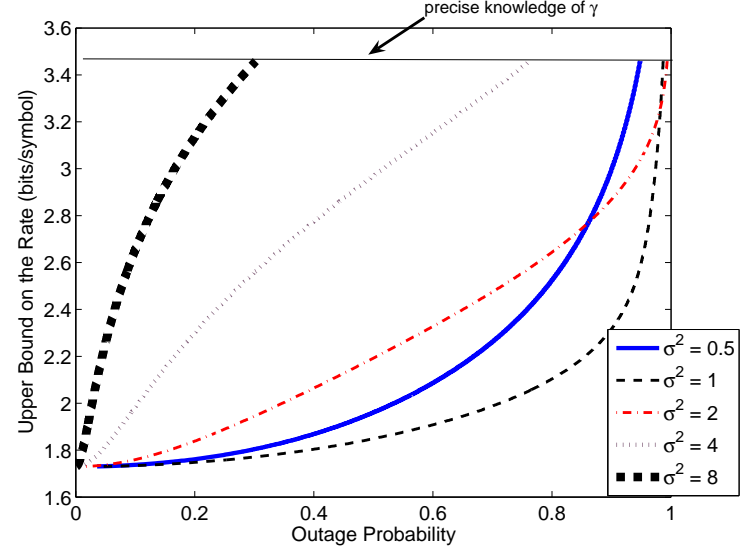


Fig. 7. The upper bound on the achievable rate vs the outage probability for varying parameter of Rayleigh distribution for  $P = Q = 10\text{db}$ . From Fig. 6, it is evident that the slowest increase in outage probability would be when the mean of the Rayleigh distribution lies in the interval. This figure demonstrates the same explicitly. The capacity of channel with no uncertainty in  $\gamma$  is  $\sim 3.46$  bits/symbol. The outage probability is, therefore, 1 above this value. We think that the bound in Fig. 6 is loose at high SIR. Therefore, the bounds for large values of  $\sigma^2$  are also loose. Observe that all the curves converge to the same value for low outage probability. This value is the same as the minimum of the upper bound in Fig. 6. This is because the bound for outage probability of zero is a bound to the achievable rate for all values of  $\gamma$ .

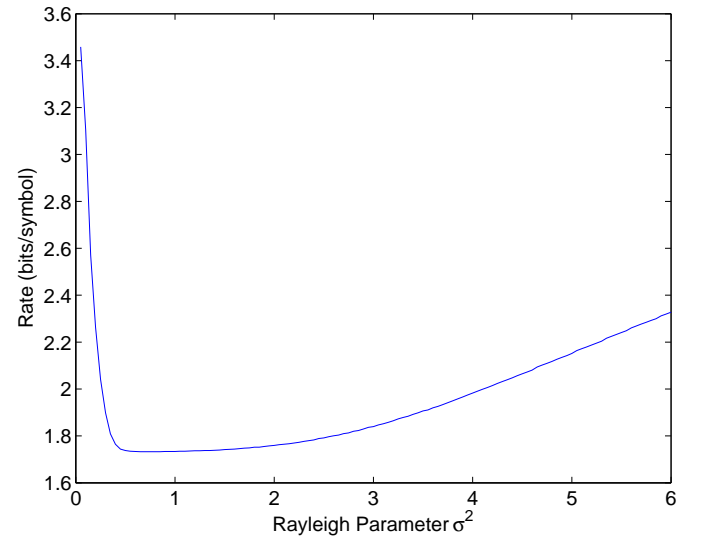


Fig. 8. The upper bound on the achievable rate vs the Rayleigh parameter  $\sigma^2$  for  $P = Q = 10\text{db}$  and an outage probability of 0.1. We think that the bound in Fig. 6 is loose at high SIR. Therefore, the bound is loose at large  $\sigma^2$ . We think that the outage capacity is a decreasing function of  $\sigma^2$ .

able rate is given by

$$R = \sup_{\alpha \in [0,1]} \log \left( \frac{P}{\bar{\mathcal{E}}(\Delta\phi)} \right) \quad (12)$$

where

$$\bar{\mathcal{E}}(\Delta\phi) = (1 - \beta(\Delta\phi))^2 P + (\alpha^2 + \beta(\Delta\phi)^2 - 2\alpha\beta(\Delta\phi) \cos(\phi))Q + \beta(\Delta\phi)^2$$

and  $\beta(\theta) := \frac{P + \alpha \cos(\theta)Q}{P + Q + 1}$

*Proof:* See Appendix II.

### B. Rates achieved by sectoring

Suppose we divide the circle of uncertainty into  $k$  sectors, each forming an angle of  $\frac{2\pi}{k}$ . Then we time-share dirty paper code with respect to the central interference vector in each sector. Intuitively, time sharing costs us a factors of  $k$  in rate. However, for the best sectors, uncertainty in phase is at worst  $\frac{\pi}{k}$ .

The rate achieved by sectoring is therefore given by

$$R = \frac{1}{k} \log \left( \frac{P}{\bar{\mathcal{E}}(\frac{\pi}{k})} \right) \quad (13)$$

To get an intuition into the achievable rates, we do some approximate calculations. The residual uncertainty is (approximately)  $Q \sin^2(\pi/k)$ . At low SIR,

$$R \approx \frac{1}{k} \log \left( 1 + \frac{P}{1 + Q \sin^2(\pi/k)} \right) \approx \frac{1}{k} \frac{Pk^2}{Q\pi^2} \quad (14)$$

if  $Q \sin^2(\pi/k) \gg 1$ . The optimal  $k$  can be determined by maximizing the rate over  $k$ . Intuitively, if  $P < 1$  and  $Q \gg 1$ , then we should keep increasing number of sectors till  $Q\pi^2/k^2 \approx 1$  (residual interference becomes comparable to 1). For this value of  $k$ , the rate obtained is  $O(\frac{P}{\sqrt{Q}})$ , which is a gain of  $\sqrt{Q}$ . If  $Q \gg P > 1$ , then we should keep increasing number of sectors till  $Q\pi^2/k^2 \approx P$  (the log approximation fails at this point). The resulting rate is  $O(\sqrt{\frac{P}{Q}})$ , a smaller advantage than for  $P < 1$ .

We plot the (low SIR) region in which the sectoring method proposed here performs better than ignoring interference as noise (see Fig. 9). Fig. 10 shows the optimal number of sectors with  $P$  for fixed interference power of 25 db. As expected, an advantage is obtained only in extremely low SIR region.

## VI. PERFORMANCE WITH FEEDBACK: PHASE ESTIMATION

In Appendix II, we derived a lower bound on rate for some uncertainty in the phase. This can happen if the

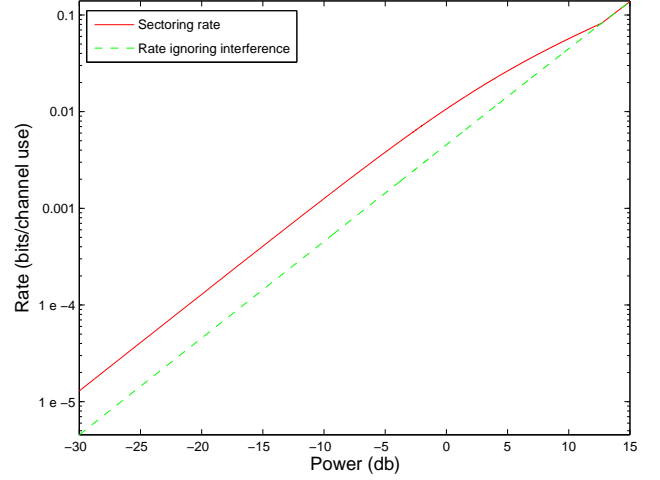


Fig. 9. Plotted on log scale, the difference between the two curves converges to a constant at low SIR. Thereby, the ratio of the two rates converges to a constant as SIR decreases.

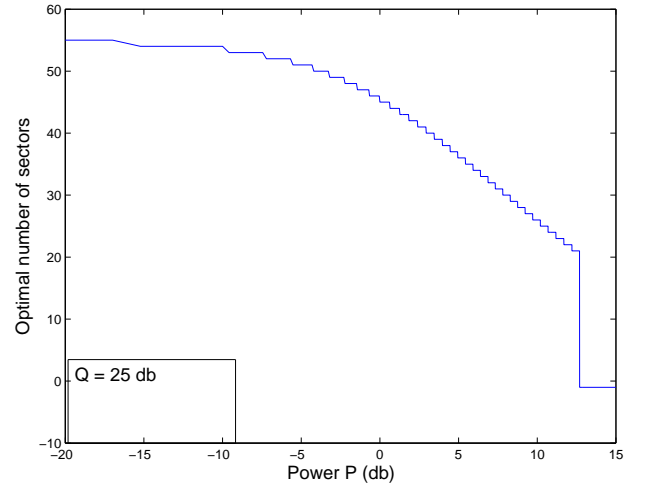


Fig. 10. Number of sectors required to achieve optimal performance decreases as SIR increases. 0 sectors corresponds to ignoring interference as noise.  $Q$  is again maintained at 25db.

transmitter gets some feedback from the receiver about phase estimate. This avoids the need to waste energy on transmitting for all possible phase sectors.

Traditionally, the problem of phase estimation is overcome by the transmission of some pilot signal, which the receiver already knows. However, it is not reasonable to assume that the primary transmitter has a pilot signal known to the secondary receiver. Therefore, the receiver cannot estimate the phase without the secondary transmitter's help, since it does not know the signal

transmitted by the interferer.

Since the interference phase uncertainty is zero-rate — it does not grow with time — it is natural to ask if the receiver can estimate the phase and convey it to the transmitter at an asymptotically vanishing cost. In this section, we provide some ideas for phase estimation, and give a scheme to combat interference in the face of residual small uncertainty.

#### A. Strategy for phase estimation

Consider the following simple strategy to estimate phase. During time instants  $\{1, 2, \dots, k\}$  (call this the “preamble” in analogy with training in traditional systems), the transmitter transfers its ‘prescient knowledge’ of the  $S_k^{k+r-1}$  interference symbols by appropriately compressing them and encoding the compressed data to send over the channel. For this transmission, the receiver treats the interference as Gaussian noise of power  $Q$ . Now, the receiver estimates the phase during the first  $r$  time instants, reducing the uncertainty of phase to  $\Delta\phi$ , which depends on channel noise as well as error introduced in quantization. The receiver then feeds back the phase estimate to the transmitter.<sup>3</sup> Finally, the transmitter encodes the actual data payload for the remainder of the time. The transmission ends at the end of a packet or the end of the coherence time, whichever comes first.

The above strategy is well suited for a packet-oriented communication scheme where the time between packets is potentially many channel coherence times and unknown to the transmitter in advance. There is not much we can do in such scenarios since the architecture dictates that every packet must be self-sufficient. The price of this self-sufficiency is that the “preamble” faces huge noise levels, since the interference is unknown while decoding it. This imposes a significant rate penalty, particularly under high SIR conditions.

The way around this problem is to assume a more continuously streaming communication model in which data will be transmitted block-after-block for many contiguous coherence times. In such cases, the block-fading model here is questionable and a Gauss-Markov model is probably more appropriate. However, inspired by [3],

<sup>3</sup>The analogy may be more accurate with the RTS/CTS messages in many wireless protocols. The “training packet” plays the role of the RTS. The receiver responds with a small “phase packet” giving the interference phase to use — in analogy with a CTS message. Only then can the main data transmission proceed. The difference is that here, this is a PHY-level consideration since it impacts waveform design rather than a MAC-level issue.

we hope that channel coherence issues are qualitatively the same in the two cases and stick with our block-fading model.

Two stages of transmission are introduced. During the initialization stage, a “preamble” as above is used. But once the transmitter acquires the phase to a certain degree of accuracy, it dirty-paper-codes the data *as well as* prescient knowledge of the first  $r$  symbols of the *next* coherence time period and sends them to the receiver. In effect, the preamble’s payload is combined with the main data payload. This is possible because unlike a pilot tone or a PRN-sequence used in a traditional preamble, we are not interested in the waveform properties of the preamble transmission. The relevant waveform for sounding out the interference channel is provided by the primary transmission — our only goal is to empower the receiver to utilize it.

At the start of next coherence time, the decoder estimates the phase using the first  $r$  time instants and feeds this information back to the transmitter. Once this has happened, the process repeats with the next data block combined with the next preamble payload. The strategy is figuratively illustrated in Fig. 11. Figure 12 (top curve) demonstrates the performance of this strategy with coherence time for  $Q = 10\text{db}$ . The underlying tension is between obtaining a good estimate for the phase to improve data rate and wasting valuable time while calculating that estimate. A higher value of interference allows for better estimation for the same dead-zone length. Therefore, the effective rate for increasing  $Q$  does not show a significant change.

The strategy is detailed and analyzed in Appendix III. The analysis is approximate because we assume the residual phase error to be small. For small values of the interference power  $Q$ , the receiver may as well ignore interference altogether, rather than spending time learning it. In that case, the phase uncertainty is  $\pi$ , and the analysis is not valid. Therefore, the analysis holds only for large  $Q$ . The accuracy can be estimated from the phase value that attains the maximum rate.

Sometimes the transmitter may not know when its next transmission would be. This situation can occur when the source generates data at random times, unknown to the transmitter in advance. In this case, the initialization step needs to be performed for each packet, and therefore, the prescient knowledge has to be decoded treating all interference as noise. After decoding prescient knowledge, the transmitter keeps silent for the ‘dead zone’, as the receiver uses the prescient knowledge to estimate the phase. The receiver then feeds the phase back, and



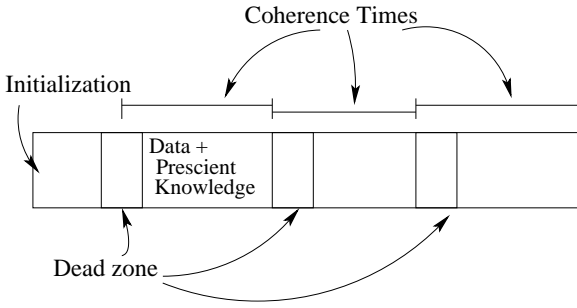


Fig. 11. Bootstrap strategy for phase estimation. The dead-zone allows for phase estimation. Transmitter sends some prescient knowledge of the interference to the receiver during the previous data block. The receiver uses this knowledge to estimate phase during the dead-zone. The process is repeated in next coherence time.

transmitter dirty-paper codes in accordance with the phase knowledge, as before.

This scheme is also analyzed at the end of Appendix III. The performance of this scheme is compared with that of the earlier strategy in Fig. 12.

## VII. CONCLUSIONS AND FUTURE WORK

We showed that the dirty paper coding strategy requires the knowledge of the interference phase at the transmitter. At high SIR, there is a significant loss in performance due to not knowing the phase. At extremely low SIR, we provided a strategy that is able to utilize the knowledge of the interferer's signal to *some* extent without having to get an estimate of the phase. We showed that if substantial gains are desired, then it is worthwhile to engineer systems with feedback where the transmitter also spends some of its power in helping the receiver estimate the interference channel.

There are many open problems here:

- The achievable schemes need to be tightened for both the case of no phase knowledge as well as when feedback is available.
- The sensitivity for multiple fading parameters beyond just phase needs to be established. Does every additional uncertainty impose its own overhead or does the case of phase uncertainty already capture all the essential effects?
- The upper bounds on the rate need to be tightened. The current bound is useless at low SINR and the current two-point uncertainty bound is also unable to adequately limit performance when the interferer is strong.
- The results need to be extended to Gauss-Markov models for fading where the variation is more continuous than the block-fading model here assumes.

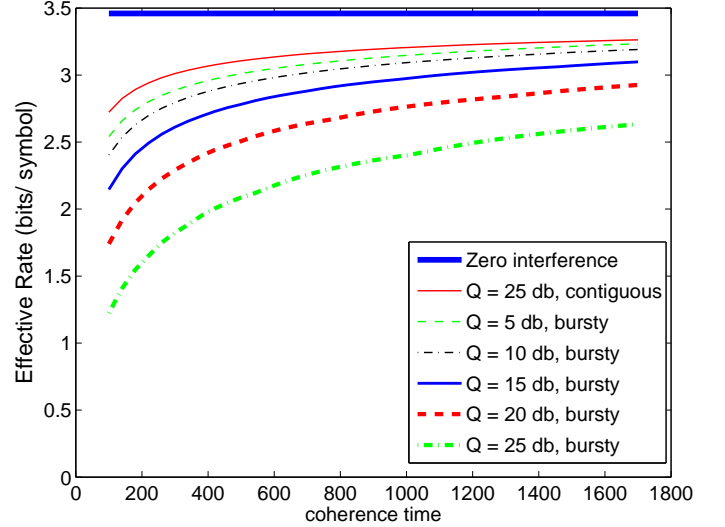


Fig. 12. Variation of effective rate achieved for  $P = 10db$  for bursty and non-bursty (contiguous time slots) transmissions. The bursty transmission results in significant loss of rate, with loss increasing for higher interference power (full phase knowledge Capacity  $\sim 3.46$  bits/symbol). For contiguous transmissions, increased coherence time allows for longer dead-zone with little loss in the rate. The variation of effective rate with  $Q$ , however, is not significant if we assume that the communication proceeds over contiguous coherent times since the data never really sees the interference. For the same relative fidelity of training quantization, a higher  $Q$  allows for better phase estimation in the dead-zone, thereby reducing the effective interference indirectly in proportion to its increase. This assumes that the higher resolution feedback message does not significantly tax the feedback link capacity.

In the end, the results in this paper are “negative results” that suggest that attempts to exploit transmitter knowledge of the interference signal realization will encounter engineering difficulties.

## APPENDIX I PROOF OF THEOREM 1

Following [13], we assume that the transmitter has phase uncertainty in the two point set  $\{0, \pi\}$ . That is, the transmitter knows that the interference vectors is either  $\mathbf{S}^n$  or  $-\mathbf{S}^n$ .

Remember that our model is

$$\begin{aligned} \mathbf{Y}_1^n &= \mathbf{X}^n + \mathbf{S}^n + \mathbf{Z}_1^n \\ \mathbf{Y}_2^n &= \mathbf{X}^n - \mathbf{S}^n + \mathbf{Z}_2^n \end{aligned} \quad (15)$$

Observe that the performance at any receiver is determined only by the marginal distribution of the noise vectors at that receiver. We are free to introduce *any* joint distribution on the noise vectors. For finding the

tightest upper bound, we assume  $\mathbf{Z}_1^n = \mathbf{Z}_2^n$ , that is, their correlation is 1.

In [13], the interference vector lies in a two-point set of statistically independent vectors. In the bounding technique there, one of the interference vectors is assumed to be zero. Using common randomness between the receiver and the transmitter, they show that this situation yields higher capacity than independent interferences of the same power. We note that for our case, the method in [13] to upper bound the capacity does not work. This is because the common randomness argument hinges on ability to generate random sequences at the transmitter and the receiver that simulate the interference ([13]). For a phase shifted interference vector, such sequence generation is not possible at the receiving end since it does not know the interference realization. A simple way of seeing this is where  $\phi = 0$ . In this case, the usual dirty-paper coding scheme is optimal. However, assuming one of the possible interference vectors to be zero would decrease the rate beyond that!

Since the transmitter is able to communicate reliably for either of these two phases, the cut-set bound and Fano's inequality [16] give us:

$$nR \leq I(W; \mathbf{Y}_i^n) + n\epsilon \quad (16)$$

And therefore,

$$nR \leq \min_i I(W; \mathbf{Y}_i^n) + n\epsilon \quad (17)$$

To avoid clutter, we drop the  $\epsilon$ 's in the following steps. From (17),

$$\begin{aligned} nR &\leq \frac{I(W; \mathbf{Y}_1^n) + I(W; \mathbf{Y}_2^n)}{2} \\ &= \frac{1}{2}(h(\mathbf{Y}_1^n) - h(\mathbf{Y}_1^n|W) + h(\mathbf{Y}_2^n) - h(\mathbf{Y}_2^n|W)) \\ &\leq \frac{1}{2}(h(\mathbf{Y}_1^n) + h(\mathbf{Y}_2^n) - h(\mathbf{Y}_1^n, \mathbf{Y}_2^n|W)) \quad (18) \end{aligned}$$

Suppose the correlation between  $X_i$  and  $S_{k,i}$  is  $\rho_{ki}$  for  $k = 1, 2$ . We observe that for random variables of given second moment, the Gaussian random variable has the maximum entropy. Thus,

$$h(\mathbf{Y}_k^n) \leq \sum_{i=1}^n \log(2\pi e(P_i + Q + 2\rho_{ki}\sqrt{P_i Q} + N)) \quad (19)$$

where  $P_i$  is the power of the  $i^{th}$  transmitted symbol,  $X_i$ .

Also,

$$\begin{aligned} h(\mathbf{Y}_1^n, \mathbf{Y}_2^n|W) &= h\left(\frac{\mathbf{Y}_1^n + \mathbf{Y}_2^n}{\sqrt{2}}, \frac{\mathbf{Y}_1^n - \mathbf{Y}_2^n}{\sqrt{2}}|W\right) \\ &= h\left(\frac{\mathbf{Y}_1^n - \mathbf{Y}_2^n}{\sqrt{2}}|W\right) + h\left(\frac{\mathbf{Y}_1^n + \mathbf{Y}_2^n}{\sqrt{2}}|W, \frac{\mathbf{Y}_1^n - \mathbf{Y}_2^n}{\sqrt{2}}\right) \quad (20) \end{aligned}$$

Since  $\mathbf{Y}_1^n - \mathbf{Y}_2^n$  is independent of  $\mathbf{X}^n$  (and hence, of  $W$ ), the first term in (20) is easy to simplify

$$\begin{aligned} h\left(\frac{\mathbf{Y}_1^n - \mathbf{Y}_2^n}{\sqrt{2}}|W\right) &= n \log(2\pi eQ(1 - \cos \pi)) \\ &= n \log(4\pi eQ) \quad (21) \end{aligned}$$

where we use  $\mathbf{Z}_1^n = \mathbf{Z}_2^n$ .

The second term can be lower bounded as follows

$$\begin{aligned} &h\left(\frac{\mathbf{Y}_1^n + \mathbf{Y}_2^n}{\sqrt{2}}|W, \frac{\mathbf{Y}_1^n - \mathbf{Y}_2^n}{\sqrt{2}}\right) \\ &\geq h\left(\frac{\mathbf{Y}_1^n + \mathbf{Y}_2^n}{\sqrt{2}}|W, \frac{\mathbf{Y}_1^n - \mathbf{Y}_2^n}{\sqrt{2}}, \mathbf{S}^n, \mathbf{X}^n\right) \quad (22) \end{aligned}$$

Notice that  $\mathbf{Y}_1^n + \mathbf{Y}_2^n = 2\mathbf{X}^n + 2\mathbf{Z}_1^n$ . Therefore,

$$\begin{aligned} h\left(\frac{\mathbf{Y}_1^n + \mathbf{Y}_2^n}{\sqrt{2}}|W, \frac{\mathbf{Y}_1^n - \mathbf{Y}_2^n}{\sqrt{2}}\right) &\geq h\left(\frac{2\mathbf{Z}_1^n}{\sqrt{2}}\right) \\ &= n \log(2\pi e * 2N) \quad (23) \end{aligned}$$

Using (20), we can now lower bound the entropy

$$\begin{aligned} h(\mathbf{Y}_1^n, \mathbf{Y}_2^n|W) &\geq n \log(4\pi eQ) \\ &\quad + n \log(4\pi eN) \quad (24) \end{aligned}$$

Now in (18), it is sufficient to bound  $h(\mathbf{Y}_1^n) + h(\mathbf{Y}_2^n)$ . Using (19),

$$\begin{aligned} &h(\mathbf{Y}_1^n) + h(\mathbf{Y}_2^n) \\ &\leq \sum_{k=1,2} \sum_i \log[2\pi e(P_i + Q + 2\rho_{ki}\sqrt{P_i Q} + N)] \\ &= \sum_i \sum_{k=1,2} \log[2\pi e(P_i + Q + 2\rho_{ki}\sqrt{P_i Q} + N)] \\ &\stackrel{(a)}{\leq} 2 \sum_i \log[2\pi e(P_i + Q + (\rho_{1i} + \rho_{2i})\sqrt{P_i Q} + N)] \\ &\stackrel{(b)}{=} 2 \sum_i \log[2\pi e(P_i + Q + N)] \\ &\stackrel{(c)}{\leq} 2n \log[2\pi e(P + Q + N)] \end{aligned}$$

Inequalities (a) and (c) follow from concavity of  $\log(\cdot)$  function. (b) follows from the fact that  $\rho_{1i} = -\rho_{2i}$ . From (24) and (25), we get the following bound on the rate:

$$R \leq \frac{1}{2} \log \left[ \frac{(P + Q + N)^2}{4QN} \right] \quad (25)$$

□

## APPENDIX II

### DERIVATION OF LOWER BOUND FOR PARTIAL PHASE KNOWLEDGE

#### A. Analysis for $\phi = \Delta\phi$

In the following, we use the techniques developed in [1, Pg. 530–532] and [17].

We first detail the encoding and decoding scheme, and then analyse them for finding achievable rates.

Generate a Gaussian  $U$ -codebook, where each element of each codeword is drawn i.i.d.  $\mathcal{CN}(0, P + \alpha^2 Q)$ . Generate  $N_U = \left(\frac{P + \alpha^2 Q}{\bar{\mathcal{E}}}\right)^{n(1-\epsilon)}$  such codewords to form the codebook. Distribute them randomly into  $2^{nR}$  bins. At the encoder, given message  $m \in \{1, \dots, 2^{\lfloor nR \rfloor}\}$  and  $\mathbf{S}^n$ , find  $\mathbf{U}^n$  that is jointly typical with  $\mathbf{S}^n$ . Send  $\mathbf{X}^n = \mathbf{U}^n - \alpha \mathbf{S}^n$ .

At the decoder, find the Linear Least Square Estimate (LLSE) of  $\mathbf{U}^n$  given  $\mathbf{Y}^n = \mathbf{X}^n + \mathbf{S}^n + \mathbf{Z}^n$ , assuming  $\phi$  is known, and its value is  $\Delta\phi$ . Let the average LLSE error in each direction be denoted by  $\bar{\mathcal{E}}$ . Construct a sphere of radius  $n(1 + \delta)\bar{\mathcal{E}}$  ( $\delta > 0$ , small) around  $\hat{\mathbf{U}}^n$ . Decode to the  $U$ -codeword in the sphere, if a unique such codeword exists. Else, declare a decoding error.

We now proceed to analyse the above scheme.

An encoding error would happen if for given  $\mathbf{S}^n$  and the message  $m$ , there is no jointly typical pair  $(\mathbf{U}^n, \mathbf{S}^n)$  in bin  $m$ .

The probability that a pair  $(\mathbf{U}^n, \mathbf{S}^n)$  is jointly typical, where each of the elements is drawn independently, is greater than  $(1 - \epsilon)2^{-nI(U;S) - n\epsilon}$  for sufficiently large  $n$ . Therefore, the expected number of jointly typical codewords in a bin is greater than  $(1 - \epsilon)2^{n\epsilon}$  if

$$N_u \times (1 - \epsilon)2^{-nI(U;S) - n\epsilon} \times 2^{-nR} \quad (26)$$

has a positive exponent. If this condition is satisfied, then as in [17], the probability of error converges to zero.

The expression  $I(U; S)$  simplifies to

$$\begin{aligned} I(U; S) &= h(U) - h(U|S) \\ &= h(U) - h(X) \\ &= \log \left( \frac{P + \alpha^2 Q}{P} \right) \end{aligned}$$

Therefore, for a positive exponent in (26),

$$R \leq \log \left( \frac{P}{\bar{\mathcal{E}}} \right) \quad (27)$$

There are two potential sources of decoding error. The transmitted codeword may not lie in  $S_U$ , or there may be another codeword  $\tilde{\mathbf{U}}^n$  in  $S_U$ .

Let us first find the LLSE error  $\bar{\mathcal{E}}$ . The LLSE estimate gives us:

$$\hat{U}_i = \frac{\text{Re}[E[U^* Y]]}{E|Y|^2} Y_i \quad (28)$$

$$= \frac{P + \alpha \cos(\Delta\phi)Q}{P + Q + 1} Y_i =: \beta(\Delta\phi) Y_i \quad (29)$$

where the LLSE error is given by

$$\begin{aligned} \bar{\mathcal{E}} &= E[|U_i - \hat{U}_i|^2] \\ &= E[|X_i + \alpha S_i - \beta(X_i + S_i e^{j\Delta\phi} + Z_i)|^2] \\ &= E[|(1 - \beta)X_i + (\alpha - \beta e^{j\Delta\phi})S_i - \beta Z_i|^2] \\ &= (1 - \beta)^2 P + (\alpha^2 + \beta^2 - 2\alpha\beta \cos(\Delta\phi))Q + \beta^2 \end{aligned}$$

The mean of the cross terms is zero, since by the dirty-paper-coding construction,  $X_i$  is approximately uncorrelated with  $S_i$ . In each of the real and the imaginary axes, the LSE is  $\bar{\mathcal{E}}/2$ .

By the weak law of large numbers, with high probability, the transmitted  $U$ -codeword falls in the sphere  $S_U$  of radius  $n(1 + \epsilon)\sqrt{\bar{\mathcal{E}}}$  centered at  $\hat{\mathbf{U}}^n$ . To find the achievable rate, we now need to bound the probability of some other codeword  $\tilde{\mathbf{U}}^n$  falling in  $S_U$ .

A Gaussian codebook of power  $P + \alpha^2 Q$  can be generated in an alternative way. Choose any radial direction randomly in the  $n$ -dimensional space. Then choose a point in this radial direction with the appropriate conditional distribution. This construction works because Gaussian distribution has uniform distribution over any radial direction.

Another such distribution is uniform distribution over a sphere of radius  $n(P + \alpha^2 Q)$  in  $n$ -dimensional. To generate a codebook with distribution uniform over the whole sphere, choose a radial direction randomly, and then choose a point in the radial direction with the appropriate conditional distribution<sup>4</sup>.

As  $n \rightarrow \infty$ , both the distributions concentrate towards the edge of the sphere. For gaussian codebook this is because of the weak law of large numbers, whereas for uniform codebook, it is because the volume in large dimensions concentrates around the surface of the sphere. For  $n$  large enough, conditioned on choosing the same radial direction, distance of a typical point on the radial direction chosen according to Gaussian codebook distribution from a typical point chosen according to uniform codebook distribution can be made smaller than  $n\epsilon$  for any  $\epsilon > 0$  for  $n$  large enough. Effectively, codewords of both the codebooks can be made arbitrarily

<sup>4</sup>which is *not* the uniform distribution in the radial direction

close to each other by choosing  $n$  large enough.

Therefore, the decoder can safely assume that the codewords were chosen uniformly. It decodes to the uniform codebook codeword, which is close to the gaussian codeword for the same message. With that assumption, the probability that some other codeword  $\hat{\mathbf{U}}^n$  falls in the sphere  $S_U$  is

$$\frac{\text{Vol}(S_U)}{\text{Vol}(U - \text{space})} = \left( \sqrt{\frac{\frac{\bar{\epsilon}}{2}}{\frac{P}{2} + \alpha^2 \frac{Q}{2}}} \right)^{2n} \quad (30)$$

Using the union bound, the probability that any other codeword falls in  $S_U$  converges to zero as long as the number of  $U$ -codewords is no greater than  $N_U = \left( \frac{P + \alpha^2 Q}{\bar{\epsilon}} \right)^{n(1-\epsilon)}$  for some  $\epsilon > 0$ .

Observe that the proof is fairly general. In particular, the proof also works for jointly gaussian  $X, S, Z$  with arbitrary correlations. The uniform codebook argument is brought in to use the volume arguments. The gaussian codebook is necessary for using typicality arguments.

### B. Analysis for $\phi < \Delta\phi$

In absence of knowledge of  $\phi$  at the receiver, we perform the same operations as for  $\phi = \Delta\phi$ . The mean squared error now is

$$\begin{aligned} \bar{\mathcal{E}}_\phi &= E[|U_i - \hat{U}_i|^2] \\ &= E[|X_i + \alpha S_i - \beta(X_i + S_i e^{j\phi} + Z_i)|^2] \\ &= E[(1 - \beta(\Delta\phi))^2 |X_i|^2 \\ &\quad + (\alpha^2 + \beta(\Delta\phi)^2 - 2\alpha\beta(\Delta\phi)\cos(\phi)) |S_i|^2 \\ &\quad + \beta^2 + \text{cross terms}] \\ &= (1 - \beta(\Delta\phi))^2 P + (\alpha^2 + \beta(\Delta\phi)^2 \\ &\quad - 2\alpha\beta(\Delta\phi)\cos(\phi)) Q + \beta(\Delta\phi)^2 \end{aligned} \quad (31)$$

The mean of the cross terms is zero, since by construction,  $X_i$  is approximately uncorrelated with  $S_i$ .

Since  $|\phi| < \Delta\phi$ , the mean square error is only smaller. Therefore, for  $|\phi| < \Delta\phi$  and for  $n$  large enough, the nearest codeword only comes nearer, for the same noise realization. Therefore, the nearest codeword still lies in the sphere. Since rest of the codewords are randomly chosen, uniformly in the sphere, using the same argument as in II-A, the probability of any another  $U$ -codeword falling into the sphere is small.

Therefore nearest neighbor decoding achieves the rate given by (27).  $\square$

## APPENDIX III

### ANALYSIS OF PROPOSED PHASE ESTIMATION SCHEME

Denote by  $\hat{\mathbf{S}}^n$  the receiver's estimate of the interference vector  $\mathbf{S}^n$ . The transmitter source-codes  $S$  before sending it to the receiver. Denote the induced distortion by  $D$ . Since it is obtained from the source code for  $\mathbf{S}^n$ , it looks like (see [18, Pg. 479])

$$\hat{\mathbf{S}}^n = \gamma \mathbf{S}^n + \zeta^n \quad (32)$$

where  $\zeta^n$  is Gaussian with variance  $\frac{(Q-D)Q}{D}$  and the scaling factor  $\gamma = \frac{Q-D}{D}$ .

For  $k$  time instants, there is no transmission (dead-zone,  $\tau$ ). During this time, the receiver estimates the phase and then sends it back to the transmitter. The feedback is modeled as instantaneous since the phase can be encoded using a very small number of bits.

The receiver correlates the received vector with the signal it receives. Let the correlation random variable at  $i^{\text{th}}$  instant be denoted by  $T_i$ . Then,

$$T_i = (\gamma S_i + \zeta_i)^* (S_i e^{j\phi} + N_i) \quad (33)$$

where  $N_i$  is the noise of unit variance.  $T_i$  can be simplified into

$$\begin{aligned} T_i &= (\gamma |S_i|^2 e^{j\phi} \\ &\quad + (\gamma S_i^* N_i + \zeta_i^* S_i e^{j\phi} + \zeta_i^* N_i)) \end{aligned} \quad (34)$$

It is clear from here that the sensitivity of the estimate is independent of what the phase  $\phi$  actually is. Therefore, without loss of generality, we assume  $\phi = 0$ . Denote the second term in (34) by  $\eta_i$ . Suppose the receiver estimates for  $\tau$  time instants. Then the error phase error is approximately  $\frac{\sum_{i=1}^{\tau} \eta_i}{\tau \gamma Q}$ .

The probability that this phase error is greater than  $\Delta\phi$  is

$$\Pr\left(\frac{\sum_{i=1}^{\tau} \eta_i}{\tau \gamma Q} > \Delta\phi\right) = \Pr\left(\frac{\sum_{i=1}^{\tau} \eta_i}{\sqrt{\tau} \gamma Q} > \sqrt{\tau} \Delta\phi\right) \quad (35)$$

Then, for large  $\tau$ , by the central limit theorem

$$\frac{\sum_{i=1}^{\tau} \eta_i}{\sqrt{\tau} \gamma Q} \sim \mathcal{N}\left(0, \frac{E[\eta_i^2]}{\gamma^2 Q^2}\right) \quad (36)$$

where  $E[\eta_i^2] = \gamma^2 Q + Q \frac{Q-D}{D} + \frac{Q-D}{D}$ . Therefore,

$$\Pr\left(\frac{\sum_{i=1}^{\tau} \eta_i}{\sqrt{\tau} \gamma Q} > \sqrt{\tau} \Delta\phi\right) \approx \mathcal{Q}\left(\frac{\sqrt{\tau} \gamma Q \Delta\phi}{\sqrt{E[\eta_i^2]}}\right) \quad (37)$$

where  $\mathcal{Q}$  is the familiar  $\mathcal{Q}$ -function. Set this probability of outage to some fixed  $\beta$  (say,  $10^{-3}$ ) that reflects the

acceptable probability of outage. The training time to reach  $\Delta\phi$  error in phase with confidence of  $1 - \beta$  is given by

$$\tau(\Delta\phi, D) = \frac{(\mathcal{Q}^{-1}(\beta))^2 E[\eta_i^2]}{\gamma^2 Q^2 \Delta\phi^2} \quad (38)$$

The number of data bits communicated  $N_{\text{data}}$  is

$$N_{\text{data}} = \max_{\tau, D} \left( (l_{\text{coh}} - \tau) C \left( P \frac{l_{\text{coh}}}{l_{\text{coh}} - \tau} \right) - \tau \log \left( \frac{Q}{D} \right) \right) \quad (39)$$

where  $l_{\text{coh}}$  is the coherence time, and  $C(P) = \log(\frac{P}{\mathcal{E}})$ . The effective rate is therefore

$$R_{\text{eff}} = \frac{N_{\text{data}}}{l_{\text{coh}}} \quad (39)$$

#### A. Performance with bursty transmissions

In this model, the transmitter no longer communicates the prescient knowledge with data bits. Instead, the initialization stage needs to be performed for each data packet, resulting in a greater decrease in the effective rate. During the initialization, the transmitter sends its prescient knowledge to the receiver with average power  $P_1$  for time  $\tau_{pk}$ . The receiver ignores the interferences completely, treating it as noise, and decodes the prescient knowledge. After initialization, the transmitter is silent for the deadzone time, and the receiver estimates phase during this time, and feeds it back to the transmitter. Then the transmitter uses this phase estimate to dirty-paper code for the given interference, and transmits a signal of average power  $P_1$ . We do not allow for different transmit powers for transmitting the prescient knowledge and the actual data.  $P_1$  is given by

$$P_1 = \frac{l_{\text{coh}}}{l_{\text{coh}} - \tau_{pk} - \tau_{\text{dead}}} \quad (40)$$

The data bits are now communicated at the rate

$$C(P_1) = \log \left( \frac{P_1}{\mathcal{E}_\phi} \right) \quad (41)$$

We now optimize over the distortion  $D$ ,  $\tau_{pk}$  and the dead-zone  $\tau_{\text{dead}}$  to obtain the largest possible number of bits communicated.

$$N_{\text{data}} = \max_{\tau_{pk}, D} \left( (l_{\text{coh}} - \tau_{\text{dead}} - \tau_{pk}) \times C(P_1) \right)$$

And the effective rate is given by

$$R_{\text{eff}} = \frac{N_{\text{data}}}{l_{\text{coh}}} \quad (42)$$

#### REFERENCES

- [1] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [2] I Telatar and DNC Tse. Capacity and mutual information of wideband multipath fading channels. *IEEE Trans. Inform. Theory*, 46(4), 2000.
- [3] L Zheng, DNC Tse, and M Medard. Channel coherence in the low snr regime. *submitted to IEEE Transactions on Information Theory*, 2005.
- [4] A. Lozano, A. M. Tulino, and Sergio Verdu. Optimum power allocation for parallel gaussian channels with arbitrary input distributions. *IEEE Trans. Inform. Theory*, 52(7):3033–3051, July 2006.
- [5] A Lapidoth, S Shamai, and M Wigger. On the capacity of fading mimo broadcast channels with imperfect side-information. *43rd Annual Allerton Conference on Communication, Control, and Computing*, Sept. 28-30, 2005.
- [6] A Lapidoth and S Shamai. "fading channels: how perfect need "perfect side information" be?". *IEEE Trans. Inform. Theory*, 48(5):1118–1134, May 2002.
- [7] N Jindal. MIMO Broadcast channels with Finite Rate Feedback. *IEEE Trans. Inform. Theory*, 52(11), November 2006.
- [8] N. Devroye, P. Mitran, and Vahid Tarokh. Achievable rates in cognitive radio channels. *IEEE Trans. Inform. Theory*, 52(5), 2006.
- [9] A. Jovicic and Pramod Viswanath. Cognitive radio: An information-theoretic perspective. *submitted to IEEE Transactions on Information Theory*, 2006.
- [10] Wei Wu, Sriram Vishwanath, and Ari Arapostathis. Capacity of Gaussian Weak Interference Channels with Degraded Message Sets. In *Conference on Information Sciences and Systems*, 2006.
- [11] S Gelfand and MS Pinsker. Coding for channels with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.
- [12] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 29(3), 1983.
- [13] Ashish Khisti, Uri Erez, Amos Lapidoth, and Gregory Wornell. Carbon copying onto dirty paper. *Submitted to IEEE Transactions on Information Theory*, available at <http://arxiv.org/abs/cs.IT/0511095>, 2006.
- [14] P Mitran, N Devroye, and V Tarokh. On compound channels with side information at the transmitter. *IEEE Trans. Inform. Theory*, 52(4):1745–1755, April 2006.
- [15] Sachin Katti, Hariharan Rahul, Wenjun Hu, Dina Katabi, Muriel Medard, and Jon Crowcroft. Xors in the air: Practical wireless network coding. In *Proceedings of 2006 ACM SigCOMM*, Pisa, Italy, 2006.
- [16] TM Cover and JA Thomas. *Elements of Information Theory*. Wiley, 1991.
- [17] TM Cover and M Chiang. Duality between channel capacity and rate distortion with two-sided state information. *IEEE Transactions on Information Theory*, vol. 48, no. 6, 2002.
- [18] RG Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.